McKean Stochastic Differential Equations and the Particle Method

Dhruv Madeka

Quantitative Research, Bloomberg L.P.

Department of Mathematics, Columbia University
Mathematics of Finance MA Program

Traditional SDE's

- Traditional Stochastic Differential Equation:

$$dX_t = b(t, X_t)\ dt + \sigma(t, X_t) \cdot dW_t, \tag{1}$$

- The pdf $p(t, \cdot)$ of X_t is solution to the Fokker-Planck PDE:

$$-\partial_t p - \sum_{i=1}^{n} \partial_i \left(b^i(t,x) p(t,x) \right) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} \left(\sum_{k=1}^{d} \sigma_k^i(t,x) \sigma_k^j(t,x,) p(t,x) \right) = 0$$

$$\lim_{t \to 0} p(t,x) = \delta(x - X_0)$$

- The PDE for p(t, .) is **linear**, so we speak of **linear** SDEs
- Uniqueness and existence proved if drift and volatility are Lipschitz or locally Lipschitz with respect to the standard $\mathbb{H}^2$ metric

McKean SDEs (Henry McKean, 1966)

- McKean SDE: $X_t \in \mathbb{R}^n$

$$dX_t = b(t, X_t, \mathbb{P}_t)\, dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \mathrm{Law}\,(X_t) \tag{2}$$

- One-factor SLV: $n = 2$, $X_t = (S_t, a_t)$
- The pdf $p(t, \cdot)$ of X_t is solution to the Fokker-Planck PDE:

$$-\partial_t p - \sum_{i=1}^{n} \partial_i \left(b^i(t, x, \mathbb{P}_t) p(t, x) \right) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} \left(\sum_{k=1}^{d} \sigma_k^i(t, x, \mathbb{P}_t) \sigma_k^j(t, x, \mathbb{P}_t) p(t, x) \right) = 0$$

$$\lim_{t \to 0} p(t, x) = \delta(x - X_0) \tag{3}$$

- It is **nonlinear** because $b^i(t, x, \mathbb{P}_t)$ and $\sigma_k^i(t, x, \mathbb{P}_t)$ depend on the unknown $p \Rightarrow$ We speak of **nonlinear** SDEs
- Uniqueness and existence proved if drift and volatility coefficients are Lipschitz-continuous functions of $\mathbb{P}_t$, with respect to the Wasserstein metric (see Sznitman, Méléard, Villani)

Wasserstein Metric

The Monge-Kantorovich Distance (or Wasserstein metric) between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined as:

$$d_{\mathrm{MK}}(\mathbb{P}_1, \mathbb{P}_2)^p = \inf_{\tau \in \mathcal{P}(\mathbb{R}^n, \mathbb{R}^n) \text{ with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} d(x,y)^p \, \tau(dx, dy) \right) \tag{4}$$

Interpretation:

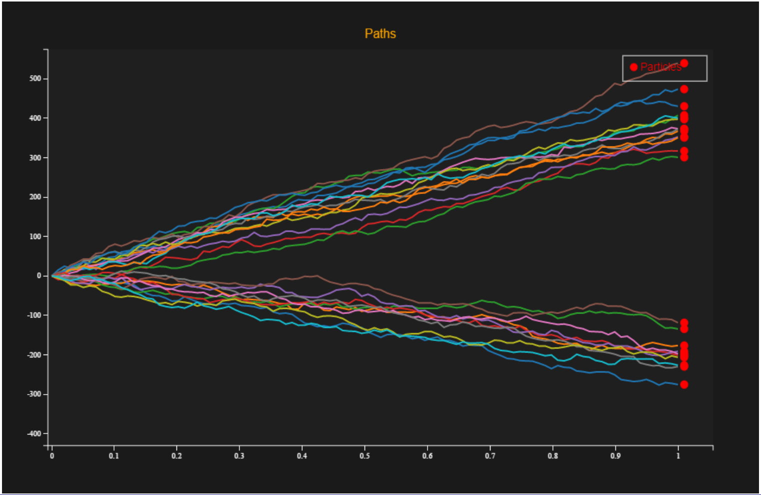
- Consider a European option on two assets $X = S_T^1$ and $Y = S_T^2$ paying $d(X,Y)^p$
- A calibrated model implies that risk-neutral distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ are marginals
- The pricing model means choosing the joint distribution τ that minimizes the option fair value:

$$\mathbb{E}^{\tau}\left[d(X,Y)^p\right]$$

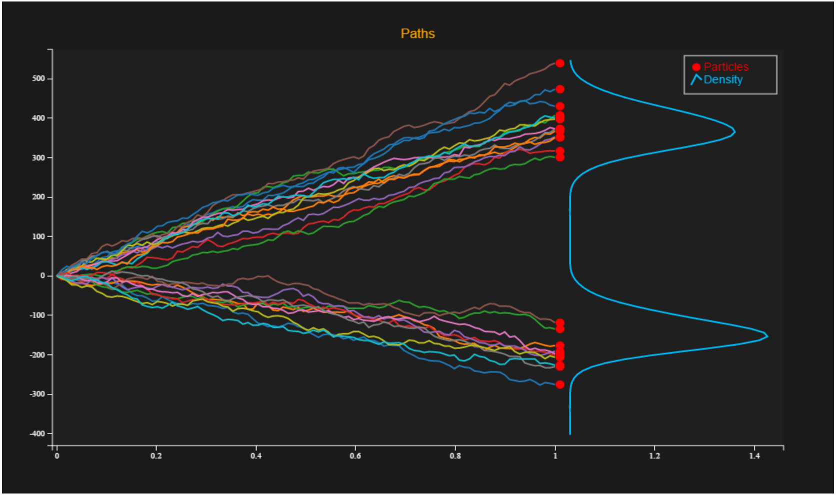
- This is precisely the Monge-Katorovich Distance

How to simulate a McKean SDE?

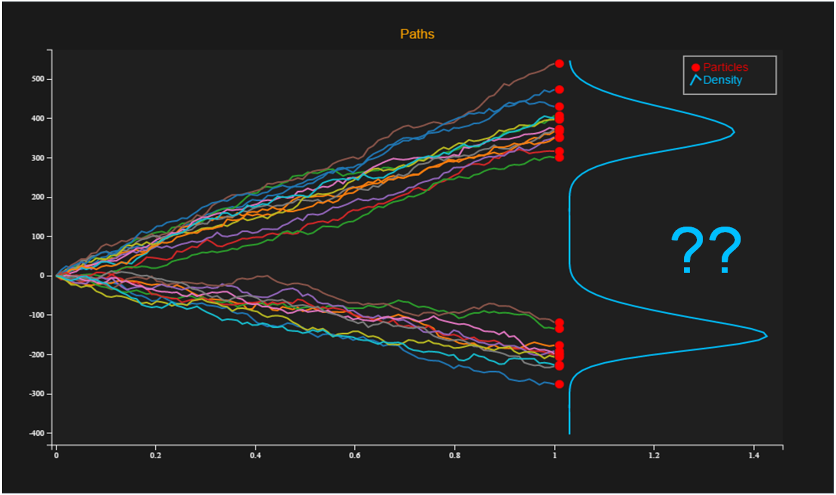
$$dX_t = b(t, X_t, \mathbb{P}_t)\ dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \text{Law}\,(X_t)$$



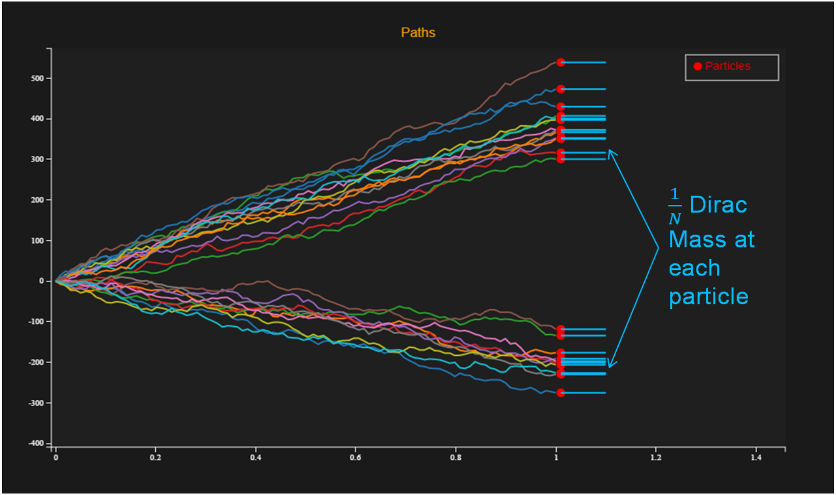
How to simulate a McKean SDE?



How to simulate a McKean SDE?



How to simulate a McKean SDE?



Particle algorithm

- Principle: replace the law $\mathbb{P}_t$ by the empirical distribution

$$\mathbb{P}_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^{i,N}}$$

where the $X_t^{i,N}$ are solution to the $(\mathbb{R}^n)^N$-dimensional **classical** SDE

$$dX_t^{i,N} = b\left(t, X_t^{i,N}, \mathbb{P}_t^N\right) dt + \sigma\left(t, X_t^{i,N}, \mathbb{P}_t^N\right) \cdot dW_t^i, \qquad \mathrm{Law}\left(X_0^{i,N}\right) = \mathbb{P}_0$$

- $\{X_t^i\}_{1 \leq i \leq N}$ = system of N **interacting particles**. The interaction with the other $N-1$ particles comes from $\mathbb{P}_t^N$
- Then (see Sznitman, Méléard, Villani) propagation of chaos $\Longrightarrow$

$$\frac{1}{N} \sum_{i=1}^{N} f(X_t^{i,N}) \xrightarrow[N \to \infty]{\mathrm{L}^1} \int_{\mathbb{R}^d} f(x) p(t,x) dx$$

- In the large N limit, the $(\mathbb{R}^n)^N$-dimensional **linear** Fokker-Planck PDE approximates the **nonlinear** low-dimensional (n-dimensional) Fokker-Planck PDE

An archetypal example: the McKean-Vlasov SDE

- For $1 \leq k \leq n$ and $1 \leq l \leq d$

$$
\begin{aligned}
b_k\left(t, x, \mathbb{P}_t\right) &= \int b_k(t,x,y)\mathbb{P}_t(dy) &= \mathbb{E}\left[b_k(t,x,X_t)\right] \\
\sigma_{kl}\left(t, x, \mathbb{P}_t\right) &= \int \sigma_{kl}(t,x,y)\mathbb{P}_t(dy) &= \mathbb{E}\left[\sigma_{kl}(t,x,X_t)\right]
\end{aligned}
$$

- Particle method:

$$dX_t^{i,N} = \left(\int b(t, X_t^{i,N}, y)d\mathbb{P}_t^N(y)\right)dt + \left(\int \sigma(t, X_t^{i,N}, y)d\mathbb{P}_t^N(y)\right)\cdot dW_t^i$$

- This is equivalent to

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} b\left(t, X_t^{i,N}, X_t^{j,N}\right)dt + \frac{1}{N}\sum_{j=1}^{N}\sigma\left(t, X_t^{i,N}, X_t^{j,N}\right)\cdot dW_t^i \tag{5}$$

- The intuition is that we use the particles of the previous step to sample X, b, σ

The Particle Method

- For convergence of the particle method, we require the chaos propagation property:
- If at $t = 0$, the $X_0^{i,N}$ are independent particles, then as $N \to \infty$, for any fixed $t > 0$, the $X_t^{i,N}$ are asymptotically independent and their empirical measure $\mathbb{P}_t^N$ converges in distribution toward the true measure $\mathbb{P}_t$.
- This means that, in the space of probabilities on the space of probabilities, the distribution of the random measure $\mathbb{P}_t^N$ converges toward a Dirac mass at the deterministic measure $\mathbb{P}_t$. Practically, it means that for all functions $\varphi \in C_b(\mathbb{R}^n)$

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int_{\mathbb{R}^n}\varphi(x)p(t,x)\,dx$$

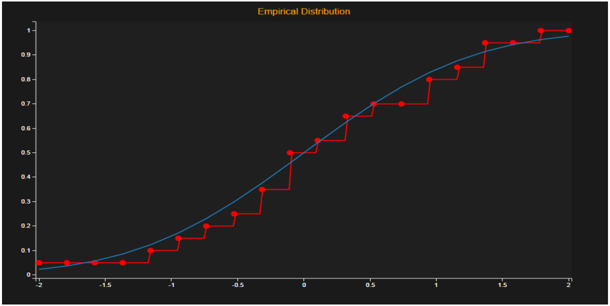
where $p(t,\cdot)$ is the fundamental solution to the nonlinear Fokker-Planck PDE (3). Hence, if propagation of chaos holds, the particle method is convergent.

Definition (Empirical measure)

Let $X_1, \ldots, X_N$ be i.i.d. random variables with law μ. The empirical measure associated to the configuration $(X_1, \ldots, X_N)$ is

$$\hat{\mu}^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_i} \tag{6}$$

Note that it is a random probability measure with expectation $\mathbb{E}_\mu[\hat{\mu}^N] = \mu$, meaning that for all events A, $\mathbb{E}_\mu[\hat{\mu}^N(A)] = \mu(A)$.



μ-chaotic distributions

Definition (μ-chaotic distribution)

Let $\{\mu^N\}_{N \in \mathbb{N}}$ be a sequence of *symmetric* probabilities on $(\mathbb{R}^n)^N$. Let μ be a probability measure on $\mathbb{R}^n$. We say that μ^N is μ-chaotic if for each integer $k \geq 1$ and for all test functions $\varphi_1, \ldots, \varphi_k \in C_b(\mathbb{R}^n)$ (i.e., for all bounded continuous functions), we have

$$\int \varphi_1(x_1) \cdots \varphi_k(x_k) \, d\mu^N(x_1, \ldots, x_N) \underset{N \to \infty}{\longrightarrow} \int \varphi_1 \, d\mu \cdots \int \varphi_k \, d\mu$$

Stated otherwise, k particles (within N) are asymptotically independent and identically distributed as $N \to \infty$ (k being fixed). We say that μ^N is chaotic if there exists μ such that μ^N is μ-chaotic.

Definition (Propagation of chaos)

Let us consider an N-dimensional SDE flow that associates to an initial probability measure μ_0^N a probability μ_t^N at time t. We say that this flow propagates the chaos if, for any initial chaotic measure μ_0^N and any $t > 0$, μ_t^N is chaotic.

Intuition

What the propagation of chaos essentially says is that as the number of particles becomes very large:

- A **fixed subset of particles** becomes independent as the sample increases
- The empirical measure converges to the law of a single particle
- It is related to the idea that in the large N-limit a linear Fokker-Planck PDE of high dimension approximates a non-linear low dimensional Fokker-Planck PDE
  - That is to say a particle among N starts behaving in a non-linear way
- It becomes apparent why this is useful for the particle method, since as N becomes large the empirical distribution depends only on the law rather than any particle

Some Theory

Theorem

Let $\{\mu^N\}_{N\in\mathbb{N}}$ be a sequence of symmetric probabilities on $(\mathbb{R}^n)^N$, and μ be a probability measure on $\mathbb{R}^n$. The following four properties are equivalent:

(i) $\{\mu^N\}_{N\in\mathbb{N}}$ *is μ-chaotic.*

(ii) *For all test functions $\varphi_1, \varphi_2 \in C_b(\mathbb{R}^n)$:*

$$\int \varphi_1(x_1)\varphi_2(x_2)\, d\mu^N(x_1, \ldots x_N) \underset{N\to\infty}{\longrightarrow} \int \varphi_1\, d\mu \int \varphi_2\, d\mu$$

(iii) *Let $X_1, \ldots, X_N$ be random variables such that $\mathrm{Law}(X_1, \ldots, X_N) = \mu^N$. Then, for all $\varphi \in C_b(\mathbb{R}^n)$,*

$$\frac{1}{N}\sum_{i=1}^{N} \varphi(X_i) \underset{N\to\infty}{\overset{\mathrm{L}^1}{\longrightarrow}} \int \varphi\, d\mu$$

(iv) *Let $\hat{\mu}^N$ be the empirical measure associated to μ^N. Then*

$$\mathbb{E}_{\mu^N}\left[\left|\int \varphi\, d\hat{\mu}^N - \int \varphi\, d\mu\right|\right] \underset{N\to\infty}{\longrightarrow} 0$$

for all $\varphi \in C_b(\mathbb{R}^n)$.

(i) $\rightarrow$ (ii) follows from the definition with k=2.
(ii) $\rightarrow$ (iii)

Proof.

We have

$$\begin{aligned}
& \mathbb{E}_{\mu^N}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\varphi(X_i) - \int \varphi\, d\mu\right)^2\right] \\
= \; & \mathbb{E}_{\mu^N}\left[\frac{1}{N^2}\sum_{i,j=1}^{N}\varphi(X_i)\varphi(X_j) - 2\int \varphi\, d\mu \frac{1}{N}\sum_{i=1}^{N}\varphi(X_i) + \left(\int \varphi\, d\mu\right)^2\right] \\
= \; & \frac{1}{N}\mathbb{E}_{\mu^N}[\varphi(X_1)^2] + \frac{N-1}{N}\mathbb{E}_{\mu^N}[\varphi(X_1)\varphi(X_2)] \\
& -2\int \varphi\, d\mu \mathbb{E}_{\mu^N}[\varphi(X_1)] + \left(\int \varphi\, d\mu\right)^2 \\
& \underset{N\to\infty}{\longrightarrow} 0 + \left(\int \varphi\, d\mu\right)^2 - 2\left(\int \varphi\, d\mu\right)^2 + \left(\int \varphi\, d\mu\right)^2 = 0
\end{aligned}$$

as from (ii) $\mathbb{E}_{\mu^N}[\varphi(X_1)\varphi(X_2)] \underset{N\to\infty}{\longrightarrow} \left(\int \varphi\, d\mu\right)^2$ and $\mathbb{E}_{\mu^N}[\varphi(X_1)] \underset{N\to\infty}{\longrightarrow} \int \varphi\, d\mu.$ □

(iii) → (iv) follows from the definition of convergence in mean (iv) → (i)

Proof.

Let $k \geq 1$ and $\varphi_1, \ldots, \varphi_k \in C_b(\mathbb{R}^n)$. We have (from the triangle inequality):

$$\left|\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)] - \int \varphi_1\, d\mu \cdots \int \varphi_k\, d\mu\right|$$
$$\leq \left|\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)] - \mathbb{E}_{\mu^N}\left[\int \varphi_1\, d\hat{\mu}^N \cdots \int \varphi_k\, d\hat{\mu}^N\right]\right|$$
$$+ \left|\mathbb{E}_{\mu^N}\left[\int \varphi_1\, d\hat{\mu}^N \cdots \int \varphi_k\, d\hat{\mu}^N\right] - \int \varphi_1\, d\mu \cdots \int \varphi_k\, d\mu\right|$$

The second term on the r.h.s. converges to zero thanks to hypothesis (iv). The first term on the r.h.s. reads

$$\left|\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)] - \frac{1}{N^k}\sum_{i_1,\ldots,i_k=1}^{N} \mathbb{E}_{\mu^N}[\varphi_1(X_{i_1})\cdots\varphi_k(X_{i_k})]\right|$$

□

Proof.

In the above sum, $\frac{N!}{(N-k)!}$ terms are such that the indices $i_1, \dots, i_k$ are all different. By symmetry, they are equal to $\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)]$ and can be added to the first term. The other terms can be bounded by M^k with $M = \sum_j ||\varphi_j||_\infty$, so the first term is bounded by

$$\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)]\left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right) + \left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right)M^k$$
$$\leq 2M^k\left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right) \underset{N\to\infty}{\longrightarrow} 0$$

$\square$

The McKean SDE propagates the chaos

Theorem

The propagation of chaos property holds for the McKean Vlasov SDE (5)

The intuition of the proof relies of a sort of "coupling method", which consists of introducing a process Y_t^i defined as:

$dY_t^i = b(Y_t^i, \mathbb{P}_t)\, dt + \sigma(Y_t^i, \mathbb{P}_t)\, dW_t^i, \qquad Y_0^i = X_0$

$b(y, \mathbb{P}_t) \equiv \int b(y,z)\mathbb{P}_t(dz)$, $\sigma(y, \mathbb{P}_t) \equiv \int \sigma(y,z)\mathbb{P}_t(dz)$, and $\mathbb{P}_t = \mathrm{Law}(X_t)$

These are standard SDEs which admit a strong solution as b and σ are Lipschitz-continuous functions. The density $q_i(t,x)$ of Y_t^i satisfies the *linear* Fokker-Planck equation

$$- \partial_t q_i(t,x) - \sum_{i=1}^{n} \partial_i \left(b^i(t,x,\mathbb{P}_t) q_i(t,x) \right) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} \left(\sum_{k=1}^{d} \sigma_k^i(t,x,\mathbb{P}_t) \sigma_k^j(t,x,\mathbb{P}_t) q_i(t,x) \right) = 0$$

with Dirac initial condition. From (3), the density $p(t,x)$ of X_t is also a solution. By uniqueness, $\mathbb{P}_t = \mathrm{Law}(Y_t^i)$.

Proposition

Let $(X^{i,N})_{1\leq i\leq N}$ be defined by (5). Then

$$\mathbb{E}[|X_t^{1,N} - Y_t^1|] \leq \frac{C(t)}{\sqrt{N}}$$

where $C(t)$ is a smooth function of time independent of N.

Proof of Theorem 5.

To prove the propagation of chaos for a McKean SDE, we use the propagation above: Let us denote by μ_t the law of the solution X_t of the McKean SDE (5). From Theorem 4 (iii), it is enough to show that for all $\varphi \in C_b(\mathbb{R}^n)$,

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int \varphi \, d\mu_t \tag{7}$$

where the $X_t^{i,N}$ are defined by (5). □

Proof of Theorem 5.

Now,

$$\mathbb{E}\left[\Big|\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) - \int \varphi\, d\mu_t\Big|\right]$$
$$\leq \mathbb{E}_{\mu^N}\left[\Big|\frac{1}{N}\sum_{i=1}^{N}\Big(\varphi(X_t^{i,N}) - \varphi(Y_t^i)\Big)\Big|\right] + \mathbb{E}_{\mu^N}\left[\Big|\frac{1}{N}\sum_{i=1}^{N}\varphi(Y_t^i) - \int \varphi\, d\mu_t\Big|\right]$$

As by construction the processes $\{Y_t^i\}_{1\leq i\leq N}$ are i.i.d. with law μ_t, the second term above goes to zero as $N \to \infty$ from the law of large numbers.
For all $\epsilon > 0$, there exists a Lipschitz-continuous function φ_ϵ such that $|\varphi - \varphi_\epsilon| \leq \epsilon$. The first term is then bounded by

$$2\epsilon + \mathbb{E}_{\mu^N}\left[\Big|\frac{1}{N}\sum_{i=1}^{N}\Big(\varphi_\epsilon(X_t^{i,N}) - \varphi_\epsilon(Y_t^i)\Big)\Big|\right]$$
$$\leq 2\epsilon + ||\varphi_\epsilon||_{\mathrm{Lip}}\mathbb{E}[|X_t^{1,N} - Y_t^1|] \leq 2\epsilon + ||\varphi_\epsilon||_{\mathrm{Lip}}\frac{C(t)}{\sqrt{N}}$$

from Proposition 21. □